



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/716,951	11/18/2003	Christopher Rohrs	025.0365.US.UTL	6464

26479 7590 10/05/2006

STRAUB & POKOTYLO
620 TINTON AVENUE
BLDG. B, 2ND FLOOR
TINTON FALLS, NJ 07724

EXAMINER

BOTTS, MICHAEL K

ART UNIT	PAPER NUMBER
----------	--------------

2176

DATE MAILED: 10/05/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 10/716,951	Applicant(s) ROHRS, CHRISTOPHER	
	Examiner Michael K. Botts	Art Unit 2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 10 July 2006.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-14, 16-48 and 50 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☒ Claim(s) 1-14, 16-22, 24-38, 40-48, and 50 is/are allowed.
- 6) ☒ Claim(s) 1-14, 16-22, 24-38, 40-48, and 50 is/are rejected.
- 7) ☒ Claim(s) 23 and 39 is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This document is a Final Office Action on the merits. This action is responsive to the following communications: Amendment, which was filed on July 10, 2006.
2. Claims 1-14, 16-48, and 50 are currently pending in the case, with claims 1, 8, 16, 17, 33, and 50 being the independent claims.
3. The drawings were objected to. Applicant has submitted new drawings obviating the grounds of the objection. Accordingly, the objection is withdrawn.
4. The Abstract was objected to. Applicant has appropriately amended the abstract. Accordingly, the objection is withdrawn.
5. Claims 15 and 49 were rejected under 25 U.S.C. 112, 2nd paragraph, and 35 U.S.C. 101, and were cancelled by the Applicant. Accordingly, the rejections are moot.
6. Claims 23 and 39 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.
7. Claims 1-14, 16-22, 24-38, 40-48, and 50 are rejected.

Claims Objections

8. **Dependent claim 32** is objected to because of the following informality: Claim 32 depends from dependent claim 31, which in turn depends from independent claim 17. Claim 32 adds the limitation "at least one record specifying at least one such word

Art Unit: 2176

as a key into the hash table” See, Claim 31, lines 2-3. The term “record” does not appear in either claim 31 or claim 17.

In Applicant’s Amendment, file July 10, 2006, applicant argued that “record” was well known to be an entry to a hash table. The Examiner disagrees with this argument. The term “record” was not known to be limited to entries of hash tables, and was known to one of ordinary skill in the art at the time of the invention to have a broader meaning, including “a data structure that is a collection of fields (elements), each with its own name and time.” See, “Microsoft Computer Dictionary,” fifth edition, Microsoft Press, 2002, definition of “record.” By the definition known to one of ordinary skill in the art at the time of the invention, a “record” was the broader data structure containing elements, not the elements themselves.

For purposes of this Office Action only, the Examiner will read the term “record” as being an entry to a hash table. However, since Applicant is using a standard term in a non-standard manner, and such use is deemed likely to cause confusion to the public in interpreting the claim, the claim is objected to on the ground of use of a non-standard term. The Applicant is required to amend the claim in the next office action to replace the term “record” with a standard term.

Appropriate correction is required.

Claims Rejections – 35 U.S.C. 112, Second Paragraph

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Where applicant acts as his or her own lexicographer to specifically define a term of a claim contrary to its ordinary meaning, the written description must clearly redefine the claim term and set forth the uncommon definition so as to put one reasonably skilled in the art on notice that the applicant intended to so redefine that claim term. *Process Control Corp. v. HydReclaim Corp.*, 190 F.3d 1350, 1357, 52 USPQ2d 1029, 1033 (Fed. Cir. 1999). The term "record" in claim 32 is used by the claim to mean "element", while the accepted meaning is "a data structure that is a collection of fields (elements), each with its own name and time." See, "Microsoft Computer Dictionary," fifth edition, Microsoft Press, 2002, definition of "record." The term is indefinite because the specification does not clearly redefine the term.

9. **Claims 1-7** are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The elements "capitalizer," "tokenizer," "processor," and "preprocessor" are defined with functions which are contradictory or mutually exclusive, specifically:

a) Independent claim 1 identifies a "capitalizer to analyze the set of words for correct capitalization." Claim 1 also identifies a separate "preprocessor to tokenize an expert of unstructured content into a set of words." The "capitalizer" is separate from the "preprocessor to tokenize." Dependent claims 2-7 inherit this definition through claim 1.

b) The disclosure identifies the capitalizer as the element that tokenizes, stating: "The capitalizer 64 tokenizes individual words." See, disclosure, page 10, lines 17-18. Further, in the same paragraph, the tokenizer is defined as a separate element, stating: "In one embodiment, individual words within the excerpt 66 are tokenized with regular expression or with a tokenizer 65." See, disclosure, page 10, lines 21-23.

c) Figure 4 illustrates a "Tokenizer 65" contained entirely within the "Capitalizer 64" element, and no other tokenizer is shown.

d) Figure 4 also illustrates a "processor 62, but does not identify a "preprocessor to tokenize" as claimed.

With the lack of clarity and apparent mutually exclusive definitions of the terms "capitalizer," "tokenizer," "processor," and "preprocessor," as identified above, one of ordinary skill in the art at the time of the invention would not be able to make or use the invention claimed.

10. In the interest of compact prosecution, the application is further examined against the prior art, as stated below, upon the assumption that the applicants may overcome the above stated rejection under 35 U.S.C. 112, second paragraph.

Claims Rejections – 35 U.S.C. 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

Art Unit: 2176

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

11. **Claims 1-14, 16-22, 24-38, 40-48, and 50 are rejected under 35 U.S.C. 102(b) as being clearly anticipated by Coden, et al. (U.S. Patent Application Publication 2002/0099744, published July 25, 2002, and issued as U.S. Patent 6,922,809 on July 26, 2005) [hereinafter "Coden"].**

Regarding **independent claim 1, as amended**, Coden teaches:

A system for providing capitalization correction for unstructured excerpts, comprising:

a preprocessor to tokenize an excerpt of unstructured content into a set of words; and

(See, Coden, Figure 1, item 50, and Figure 2, and paragraph [0035], teaching a preprocessor to output character data to other parts of the system, including titles, abbreviations, single words and phrases.)

a capitalizer to analyze the set of words for correct capitalization, comprising:

an evaluator to evaluate individual characters constituting at least one such word in the set of words; and

(See, Coden, paragraphs [0039]-[0053], teaching the evaluation of sentences and abbreviations, including titles and middle initials in proper names.)

a filter to skip the at least one such word if determined to be of a predefined type such that the capitalizer does not capitalize the at least one such word.

(See, Coden, Figure 12, and paragraphs [000017], [0067]-[0071], [0093], and claims 10 and 19, teaching filtering words of predefined types. The phrase: “*such that the capitalizer does not capitalize the at least one such word*” is taught as one of the variant phrases in the phrase dictionary, filter. See, Coden, paragraph [0067].)

Regarding **dependent claim 2, as amended**, Coden teaches:

A system according to Claim 1, further comprising:

a document title capitalizer to provide one or more of the words with an initial letter in uppercase and each remaining letter in lowercase.

(See, Coden, paragraph [0042], teaching a titles dictionary or simply a titles list. It is inherent in the specification of something with a title and words that it is a document.)

Regarding **dependent claim 3**, Coden teaches:

A system according to Claim 1, further comprising:

a sentence capitalizer to provide only an initial such word with an initial letter in uppercase and each remaining letter in lowercase.

(See, Coden, paragraph [0039], teaching automatic capitalization of a sentence.)

Regarding **dependent claim 4, as amended**, Coden teaches:

A system according to Claim 1 wherein the predefined type is one of (a) a word comprising a number, (B) a word including no vowels, and (c) a word not occurring at a start of a phrase and constituting at least one of an article, conjunction, preposition.

(See, Coden, Figure 6, item 325, and paragraph [0083], teaching parsing and determination whether the parsed string begins with a character or a number. See also, Coden, paragraph [0043], teaching parsed words including numbers and words consisting entirely of consonants. It is inherent in the Coden capitalization system that numbers skipped and not capitalized because numbers are incapable of being capitalized.)

Regarding **dependent claim 5**, Coden teaches:

*A system according to Claim 1, further comprising:
a lexicon comprising one or more reference words with at least one reference word defining a form of capitalization for the reference word;
a matcher to match the at least one such word against the reference words, the evaluator skipping each such word if a matching reference word is found.*

(See, Coden, paragraphs [0041]-[0071], teaching titles dictionary, abbreviations dictionary, heuristic processing, capitalization dictionary, named entity recognizer, and

Art Unit: 2176

the phrase dictionary. All such dictionaries are used for word comparison and the words are not further processed if they match the appropriate capitalization as found in the dictionaries.)

Regarding **dependent claim 6**, Coden teaches:

A system according to Claim 1, further comprising:

a proper noun capitalizer to provide the individual letters in each such word comprising a noun with no vowels in uppercase.

(See, Coden, paragraphs [0042]-[0043], teaching capitalization of titles and words consisting entirely of consonants.)

Regarding **dependent claim 7, as amended**, Coden teaches:

A system according to Claim 1, wherein the preprocessor tokenizes the excerpt into the one or more words and one or more punctuation marks.

(See, Coden, paragraph [0035], teaching a preprocessor and processing subsystems, including a punctuation subsystem, a singles subsystem, and a phrase processing subsystem.)

Regarding **independent claim 8, as amended**, claim 8 incorporates substantially similar subject matter as claimed in claim 1 and is rejected along the same rationale.

Regarding **dependent claims 9 and 10**, claims 9 and 10 incorporate substantially similar subject matter as claimed in claim 2 and are rejected along the same rationale.

Regarding **dependent claims 11-14**, claims 11-14 incorporate substantially similar subject matter as claimed in claims 4-7, respectively, and are rejected along the same rationale.

Regarding **independent claim 16**, claim 16 incorporates substantially similar subject matter as claimed in claim 1 and is rejected along the same rationale.

Regarding **independent claim 17, as amended**, Coden teaches:

A system for building a lexicon for use in capitalization correction for unstructured excerpts, comprising:

(A "lexicon" is read as synonymous with the term "dictionary." See, The American Heritage College Dictionary, Fourth Edition, Houghton Mifflin, 2002, definition of "lexicon.")

a ripper assembling a list of word sets from unstructured content, each word set comprising a word and at least one variation on capitalization;

(A "ripper" is defined in the disclosure as follows: "The ripper 34 retrieves excerpts from the text corpus 38 and tokenizes the excerpts into individual tokens from which the individual words, capitalization variations, and frequencies of occurrence are

Art Unit: 2176

determined.” See, disclosure, page 8, lines 15-17. The same function as a “ripper” is performed in the invention in Coden as a preprocess, which is defined in Coden as follows: “The output of the source 1 is connected to a preprocessor 50, described in FIG. 2, which outputs preprocessed character data to the other component parts and subsystems of the capitalization recovery system 10. These subsystems include a title processing subsystem 100, an abbreviations processing subsystem 200, a punctuation processing subsystem 300, a singles or singleton processing subsystem 500 and a phrase processing subsystem 800.” See, Coden, paragraph [0035]. A “ripper” in the application is the same element, or performs the same function as a “preprocessor” in Coden.)

an aggregator aggregating each word set, comprising:

(It is noted that the “subsystems” in Coden, identified above, are the same elements or perform the same functions as the “aggregator” element in the application. The application identifies the function of the “aggregator” as generating the lexicon. See, disclosure, page 8, lines 22-23. Similarly, the “subsystems” generate the various dictionaries in Coden. See, Coden, Figures 1-12, and paragraphs [0035]-[0093].)

an analyzer identifying at least one word set comprising significant statistics; and

(“Significant statistics” in the application include, for example, “four or more occurrences” of capitalization variation in the “text corpus” or “other forms of statistical and metrics of significance” in occurrences of capitalization variation. See, disclosure, page 8, lines 22-31. Identically, Coden tracks significant statistics, which directly impact

Art Unit: 2176

the elements in the dictionaries. See, Coden, generally, paragraphs [0055]-[0093], and specifically, paragraphs [0055]-[0063] teaching the capitalization frequency dictionary.)

a non-standard capitalization selector selecting at least two such variations within the identified word set having a non-standard capitalization, and adding the at least two such variations to the lexicon.

(See, Coden, generally, Figures 1-12, and paragraphs [0035]-[0093], and specifically, paragraph [0067] teaching adding new words to the dictionaries.

Specifically, Coden teaches that for every word variant, if its capitalization probability (according to the capitalization dictionary) is greater than 0.5, then it is added to the singles dictionary as a singleton. It is inherent from this teaching that two or more variations may be added to the lexicon, dictionary. See, Coden, paragraph [0067].)

Regarding **dependent claim 18**, Coden teaches:

A system according to Claim 17, further comprising:

a tokenizer tokenizing the excerpt into the one or more words and one or more punctuation marks.

(See, Coden, paragraph [0035].)

Regarding **dependent claim 19**, Coden teaches:

A system according to Claim 18, wherein hyphenated words are split into a plurality of the words.

Art Unit: 2176

(See, Coden, paragraph [0086], stating: "if the word is hyphenated, each word by itself is looked up in the singles dictionary 15A and the same rules as just described apply to each of the words separately before recombining them with a hyphen.")

Regarding **dependent claim 20**, Coden teaches:

A system according to Claim 17, wherein at least one variation appearing at the start of a sentence is skipped.

(See, Coden, paragraph [0040], teaching that a prior period may or may not signal the beginning of a sentence to follow and capitalization may or may not be appropriate.)

Regarding **dependent claim 21**, Coden teaches:

A system according to Claim 20, wherein the non-standard capitalization comprises the at least one variation occurring in an excerpt having fewer than half of individual letters provided in uppercase.

(See, Coden, paragraph [0064], teaching handling common words that may also be used as surnames.)

Regarding **dependent claim 24**, Coden teaches:

A system according to Claim 17, wherein the non-standard capitalization comprises the at least one variation having any individual letter other than the first individual letter provided in uppercase.

Art Unit: 2176

(See, Coden, paragraphs [0067] and [0070], teaching that "usual capitalization" means having only the first letter of a word capitalized, and that singletons in the singles dictionary may have preferred unusual capitalization forms – inherently other than the first letter.)

Regarding **dependent claim 25, as amended**, Coden teaches:

A system according to Claim 17, further comprising:

a standard capitalization selector selecting at least two such variations within the identified word set having a standard capitalization, and adding the at least two such variations to the lexicon.

(See, Coden, generally, Figures 1-12, and paragraphs [0035]-[0093], and specifically, paragraph [0067] teaching adding new words to the dictionaries. Specifically, Coden teaches that for every word variant, if its capitalization probability (according to the capitalization dictionary) is greater than 0.5, then it is added to the singles dictionary as a singleton. It is inherent from this teaching that two or more variations may be added to the lexicon, dictionary. See, Coden, paragraph [0067].)

Regarding **dependent claim 26, as amended**, Coden teaches:

A system according to Claim 17, further comprising:

a validator applying implicit rules for capitalization, and skipping each of the at least two variations subject to at least one such implicit rule.

Art Unit: 2176

(It is noted that the disclosure defines "implicit rules" in a non-limiting way, including, as stated: "The rules 41 include, by way of non-exclusive example, ignoring words that contain a number, have no vowels, or which constitute an article, conjunction, or preposition shorter than five characters and not appearing at the start of a phrase."

See, Disclosure, page 10, lines 25-19.

Coden teaches that for every word variant, if its capitalization probability (according to the capitalization dictionary) is greater than 0.5, then it is added to the singles dictionary as a singleton. It is inherent from this teaching that two or more variations may be added to the lexicon, dictionary. See, Coden, paragraph [0067].

See also, Coden, Figure 5, and paragraph [0043], teaching heuristic processing for capitalizing words consisting entirely of consonants followed by a period.)

Regarding **dependent claim 27, as amended**, Coden teaches:

A system according to Claim 26, wherein the implicit rules comprise skipping each of the at least two variations based on position within a sentence or phrase.

(See, Coden, paragraphs [0043]-[0052], teaching rules for capitalization when a period may indicate an abbreviation or the end of a sentence. Also, Coden teaches that for every word variant, if its capitalization probability (according to the capitalization dictionary) is greater than 0.5, then it is added to the singles dictionary as a singleton. It is inherent from this teaching that two or more variations may be added to the lexicon, dictionary. See, Coden, paragraph [0067].)

Regarding **dependent claim 28, as amended**, Coden teaches:

A system according to Claim 26, wherein the implicit rules comprise at least one of (A) a number, (B) having no vowels, and (C) constituting at least one of an article, conjunction and preposition.

(See, Coden, paragraph [0043], specifically teaching words consisting entirely of consonants, and see Coden, generally, Figures 1-12, and paragraphs [0035]-[0093], and specifically, paragraph [0067] teaching rules for adding new words to the dictionaries.)

Regarding **dependent claim 29, as amended**, Coden teaches:

A system according to Claim 26, wherein the implicit rules comprise normalizing a number of occurrences for each of the at least two variations using at least one of a normalizing function and relative to a source of the each of the at least two variations.

(It is again noted that the “normalizer” and the act of “normalizing” is disclosed as using the filter to protect the lexicon from being influenced by a large body of improperly capitalized words, such as might occur if the corpus of text was drawn from the web. See, disclosure, page 12, line 29 through page 13, line3.

Coden teaches that for every word variant, if its capitalization probability (according to the capitalization dictionary) is greater than 0.5, then it is added to the

Art Unit: 2176

singles dictionary as a singleton. It is inherent from this teaching that two or more variations may be added to the lexicon, dictionary. See, Coden, paragraph [0067].

See, Coden, paragraph [0017], teaching a filter to eliminate items with a high likelihood of being erroneous. See generally, Coden, paragraphs [0009]-[0021], teaching prior art and the invention to establish and protect dictionaries from infrequently occurring and erroneous entries.)

Regarding **dependent claim 30, as amended, as amended**, Coden teaches:

A system according to Claim 26, wherein the implicit rules comprise accommodating multiple forms of capitalization for each of the at least two variations by annotating each capitalization form with a frequency count and skipping those of the each of the at least two variations occurring infrequently.

(See, Coden, paragraph [0017], teaching a filter to eliminate items with a high likelihood of being erroneous. See generally, Coden, paragraphs [0009]-[0021], teaching prior art and the invention to establish and protect dictionaries from infrequently occurring and erroneous entries.

Also, Coden teaches that for every word variant, if its capitalization probability (according to the capitalization dictionary) is greater than 0.5, then it is added to the singles dictionary as a singleton. It is inherent from this teaching that two or more variations may be added to the lexicon, dictionary. See, Coden, paragraph [0067].)

Regarding **dependent claim 31**, Coden teaches:

*A system according to Claim 17, further comprising:
a hash table maintaining the lexicon.*

(See, Coden, paragraphs [0090] and [0091], teaching the use of hash tables to maintain entries to dictionaries.)

Regarding **dependent claim 32**, Coden teaches:

*A system according to Claim 31, further comprising:
at least one record specifying at least one such word as a key into the
hash table, and associating at least one such variation within the word set as a
preferred capitalization.*

(As discussed above and as disclosed, that the purpose of the hash table in Coden is to maintain entries to dictionaries. In Coden, the dictionaries are the same as the lexicons identified in the application. Identifying a key word in a hash table to set the capitalization of a word is the purpose of the table and the invention of Coden. It is inherent in Coden to set preferred capitalization of at least one word in the hash table. See, Coden, generally, and see specifically paragraphs [0090]-[0091].

Regarding **claims 33-37 and 40-48**, claims 33-37 and 40-48 incorporate substantially similar subject matter as claimed in claims 17-21 and 24-32, respectively, and are rejected along the same rationale.

Regarding **independent claim 50**, claim 50 incorporate substantially similar subject matter as claimed in claim 17 and is rejected along the same rationale.

12. It is noted that any citations to specific, pages, columns, lines, or figures in the prior art references and any interpretation of the references should not be considered to be limiting in any way. A reference is relevant for all it contains and may be relied upon for all that it would have reasonably suggested to one having ordinary skill in the art. See, MPEP 2123.

Claims Rejection – 35 U.S.C. 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

13. **Claims 22 and 38 are rejected under 35 U.S.C. 103(a) as being unpatentable over Coden, et al. (U.S. Patent Application Publication 2002/0099744, published July 25, 2002, and issued as U.S. Patent 6,922,809 on July 26, 2005) [hereinafter "Coden"], in view of Katariya, et al. (U.S. Patent 6,549,897 B1, issued April 15, 2003) [hereinafter "Katariya"].**

Regarding **dependent claim 22,as amended**, Coden in view of Katariya teaches:

A system according to Claim 17, further comprising:

a normalizer normalizing a plurality of the words extracted relative to a source of the unstructured excerpt.

(It is noted that the “normalizer” and the act of “normalizing” is disclosed as using the filter to protect the lexicon from being influenced by a large body of improperly capitalized words, such as might occur if the corpus of text was drawn from the web. See, disclosure, page 12, line 29 through page 13, line3.

Coden teaches the invention of claim 17, but does not expressly teach a normalizer normalizing a plurality of the words extracted relative to a source of the unstructured excerpt.

Katariya teaches a method for normalizing word counts, a plurality of words, such as would be found in a Web environment where “false promotion” or “spamming” by advertisers inaccurately inflates a word count. See, Katariya, figures 1-4, and col. 1, line 19 through col. 18, line 52.

Coden and Katariya are combinable in that they both involve the art of evaluation of frequency and variants of words and phrases.

It would have been obvious to one of ordinary skill in the art at the time of the invention to have combined to teachings of Coden and Katariya.

The suggestion or motivation for the combination is that Katariya merely expands the capability of the method of Coden, wherein Coden can examine a large body of

Art Unit: 2176

documents with improved accuracy by normalizing a potentially confounding data situation, such as spamming. Katariya clearly solves a potential problem with Coden when Coden is used on extremely large and diverse databases, such as the Internet.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to have combined the teachings of Coden and Katariya to result in the invention specified in claim 22.)

Regarding **dependent claim 38**, claim 38 incorporate substantially similar subject matter as claimed in claim 22 and is rejected along the same rationale.

14. It is noted that any citations to specific, pages, columns, lines, or figures in the prior art references and any interpretation of the references should not be considered to be limiting in any way. A reference is relevant for all it contains and may be relied upon for all that it would have reasonably suggested to one having ordinary skill in the art. See, MPEP 2123.

Allowable Subject Matter

Claims 23 and 39 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The closest prior art is Coden, which fails to teach or suggest the combination of claims 17 and 23 or the combination of claims 33 and 39, in so far as the combined

Art Unit: 2176

claims specify that the set of non-standard capitalization selector select only words where there are at least two word variations s that are within the set of statistically significant words wherein the statistically significant words have at least four occurrences of at least one non-standard capitalization variation within a word set.

Response to Arguments

Applicants' arguments filed July 10, 2006 have been fully considered, but they are not persuasive.

Regarding objection to claim 32:

Applicants argue that "record" was a well understood term by one of ordinary skill in the art at the time of the invention to mean an entry to a hash table. See, Remarks/Arguments, page 22.

The Examiner disagrees.

The term "record" was not known to be limited to entries of hash tables, and was known to one of ordinary skill in the art at the time of the invention to have a broader meaning, including "a data structure that is a collection of fields (elements), each with its own name and time." See, "Microsoft Computer Dictionary," fifth edition, Microsoft Press, 2002, definition of "record." By the definition known to one of ordinary skill in the art at the time of the invention, a "record" was the broader data structure containing elements, not the elements themselves. Accordingly, the objection remains.

Based on the Applicant's stated definition of the term "record" and noting that it is

Art Unit: 2176

an inconsistent use of a term with an accepted meaning, a rejection under 35 U.S.C.

112, 2nd paragraph is also made above.

Regarding rejections of claims 1-7 under 35 U.S.C. 112, 2nd paragraph:

Applicant argues that the terms are not indefinite, citing that the lexicon builder includes a tokenizer and is a preprocessor.

The Examiner disagrees.

For the reasons re-stated in the rejection above, the rejections are maintained.

Regarding rejections of independent claims 1, 8, and 16:

Applicant argues that Coden does not teach skipping at least one word of a set if it is determined to be of a predefined type such that the capitalizer does not capitalize the at least one such word. See, Applicants Remarks/Amendment, pages 27-28.

The Examiner disagrees.

Coden teaches predetermined types as entries in a phrase dictionary wherein the phrase may include words that are not capitalized. See, Coden, paragraphs [0067]-[0071].

Regarding rejections of dependent claim 2:

Applicant argues that Coden does not teach "a document title capitalizer." See, Applicants Remarks/Amendment, page 28.

The Examiner disagrees.

The term “document title” is merely a preferred use for the invention. Whether the invention is used to capitalize a document title or the document itself does not distinguish over the prior art. Therefore, the specification of a “document title” is read as non-functional descriptive language which does not change the specification of a “capitalizer.”

In addition, a document title is a phrase. Coden teaches capitalizing phrases. See, Coden, paragraphs [0067]-[0071].

Regarding rejections of dependent claims 4 and 11:

Applicant argues that Coden does not define the “predefined types” of words skipped. See, Applicants Remarks/Amendment, page 28.

The Examiner disagrees.

Coden teaches a phrase and singleton dictionary that includes variants on capitalization, including words which are to be skipped for capitalization. See, Coden, paragraphs [0067]-[0071].

Regarding rejections of independent claims 17, 33, and 50:

Applicant argues that Coden does not teach or suggest “an act or element which selects at least two capitalization variations within the identified word set having a non-standard capitalization, and adding the at least two such variations to a lexicon. See, Applicants Remarks/Amendment, page 29.

The Examiner disagrees.

See, Coden, paragraph [0067], teaching that for every concept, including phrases, that occur in at least three documents are stored in their multi-word variants. Therefore, for each canonical form of a phrase, there are at least two other variants of capitalization added to the phrase dictionary, lexicon.

Regarding rejections of dependent claims 22 and 38:

Applicant argues that Cohen does not teach or suggest “an act or element which normalizes a plurality of words extracted relative to a source of the unstructured content. See, Applicants Remarks/Amendment, page 29.

The Examiner disagrees.

It is noted that the “normalizer” and the act of “normalizing” is disclosed as using the filter to protect the lexicon from being influenced by a large body of improperly capitalized words, such as might occur if the corpus of text was drawn from the web. See, disclosure, page 12, line 29 through page 13, line 3.

Coden teaches the invention of claim 17, but does not expressly teach a normalizer normalizing a plurality of the words extracted relative to a source of the unstructured excerpt.

Katariya teaches a method for normalizing word counts, a plurality of words, such as would be found in a Web environment where “false promotion” or “spamming” by advertisers inaccurately inflates a word count. See, Katariya, figures 1-4, and col. 1, line 19 through col. 18, line 52.

Art Unit: 2176

Coden and Katariya are combinable in that they both involve the art of evaluation of frequency and variants of words and phrases.

It would have been obvious to one of ordinary skill in the art at the time of the invention to have combined the teachings of Coden and Katariya.

The suggestion or motivation for the combination is that Katariya merely expands the capability of the method of Coden, wherein Coden can examine a large body of documents with improved accuracy by normalizing a potentially confounding data situation, such as spamming. Katariya clearly solves a potential problem with Coden when Coden is used on extremely large and diverse databases, such as the Internet.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to have combined the teachings of Coden and Katariya to result in the invention specified in claim 22.

Regarding rejections of dependent claims 23 and 39:

Applicant argues that Coden fails to teach or suggest "that the set comprising significant statistics comprises only non-standard capitalization variations having at least four occurrences of at least one such variation within a word set." See, Applicants Remarks/Amendment, page 30.

The Examiner agrees. See allowable subject matter above.

Regarding rejections of dependent claims 28, 29, 44, and 45:

Applicant argues that Coden fails to teach or suggest “implicit rules for capitalization comprise at least one of a number, having no vowels, and constituting at least one of an article, conjunction and preposition.” Applicant argues further that separate teachings of the implicit rules does not teach the elements in the claim. See, Applicants Remarks/Amendment, pages 30-31.

The Examiner disagrees.

See, Coden, paragraph [0043], specifically teaching words consisting entirely of consonant – implicitly “no vowels” – or a number. Therefore, Coden expressly teaches capitalization rules for two of the three items listed in the claims.

Regarding rejections of dependent claims 30 and 46:

Applicant argues Coden fails to teach or suggest “annotating each capitalization form with a frequency count and skipping those variations that occur infrequently. See, Applicants Remarks/Amendment, page 32.

The Examiner disagrees.

See, Coden, paragraph [0067], teaching that an annotation of each capitalization form is kept, by stating that those forms with a capitalization probability greater than 0.5 will be added to the singles dictionary. This implicitly teaches that the capitalization forms are kept with annotations of capitalization probability.

Conclusion

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** for the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael K. Botts whose telephone number is 571-272-5533. The examiner can normally be reached on Monday through Friday 8:00-4:00 EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather Herndon can be reached on 571-272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the

Art Unit: 2176

Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free).

MKB/mkb


Heather R. Herndon
Supervisory Patent Examiner
Technology Center 2100